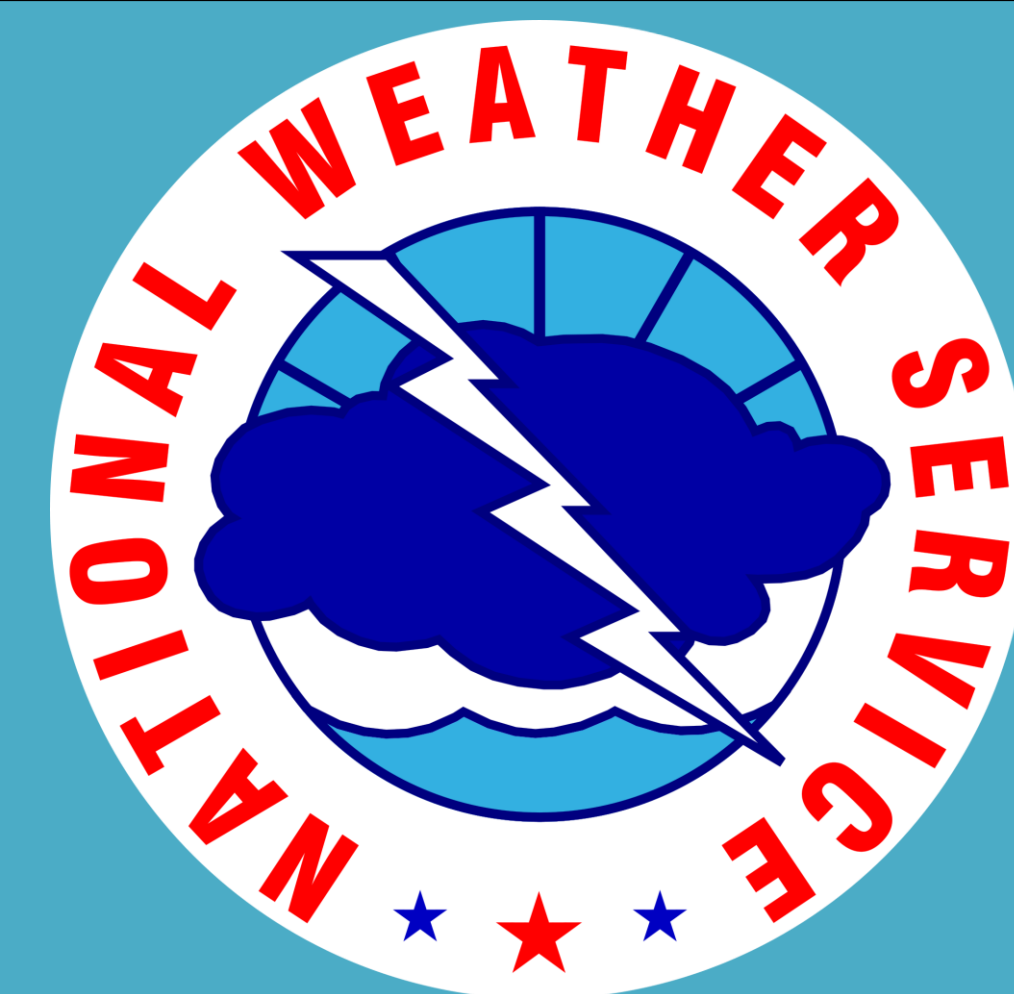


# Separation of Wind Wave Systems Using K-Means Clustering

André van der Westhuisen<sup>1</sup>

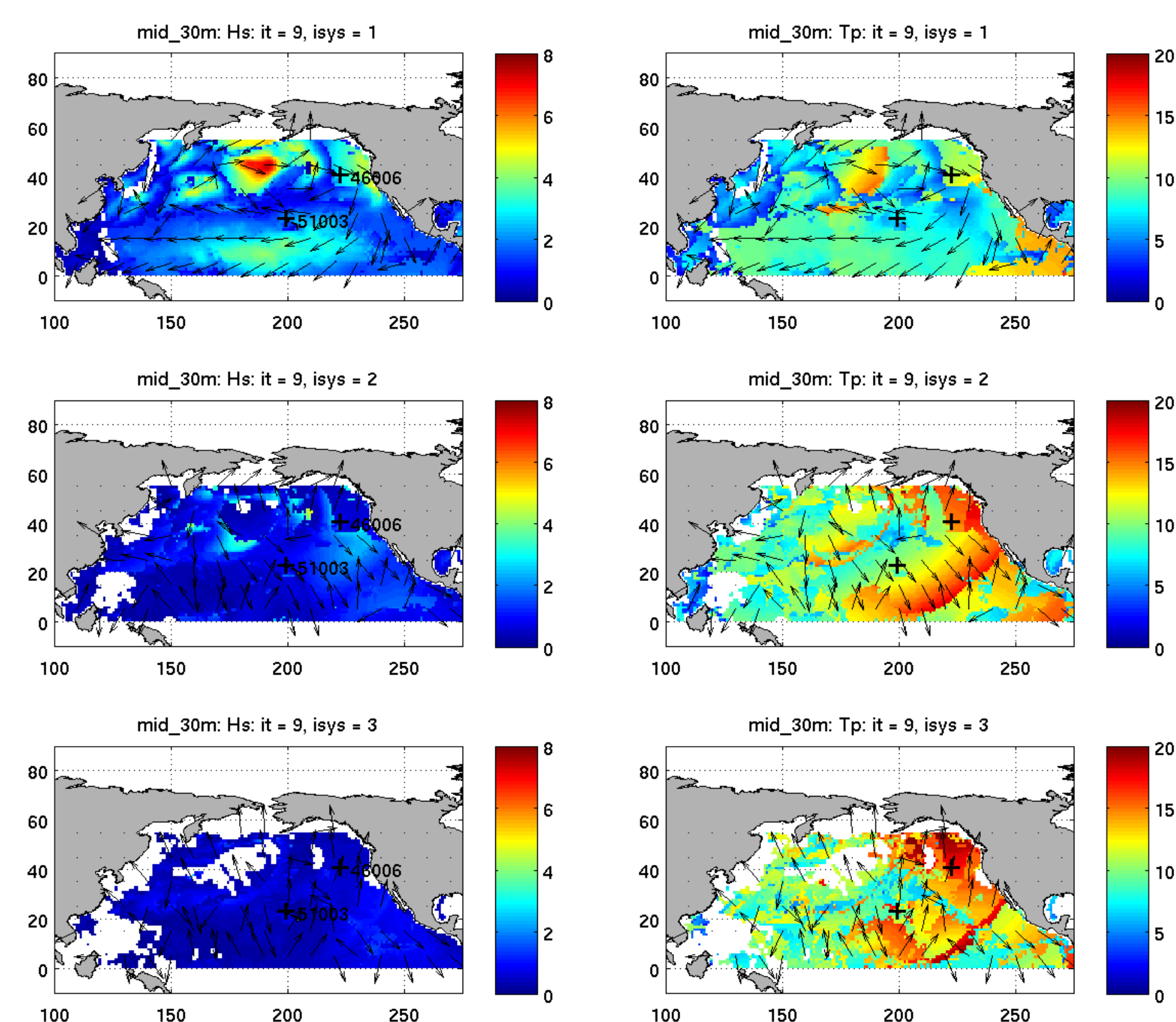
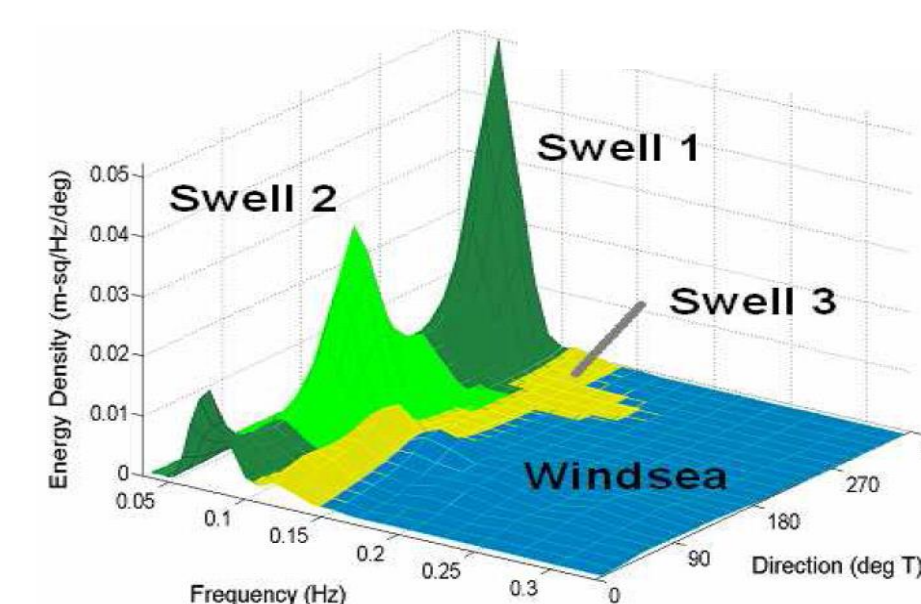
<sup>1</sup>IMSG at NWS/NCEP/EMC, College Park, MD, United States



## Background

Third-generation spectral wind wave models such as WAVEWATCH III (Tolman et al. 2016) produce output with a very large number of degrees of freedom (100M to 1B per time step on a typical model grid). To reduce this large amount of information, while retaining details of complex wave fields, wave spectrum partitioning algorithms have been developed to group significant wave components such as swells and wind sea.

This partitioned model output is increasingly being applied to provide targeted forecasting for specific marine activities (e.g. wave response of large commercial ships alongside conditions for small recreational craft). However, these partitioning algorithms operate locally in geographical space, independently at each grid point. As such, the coherence of the derived swell and wind sea partitions in geographical space and time cannot be guaranteed. In the present study, we propose an unsupervised machine learning approach for combining these independently-computed wave partitions into spatially and temporally consistent wave systems. This task is cast as a clustering problem, which is solved using the K-Means algorithm from Python's Scikit-Learn package.

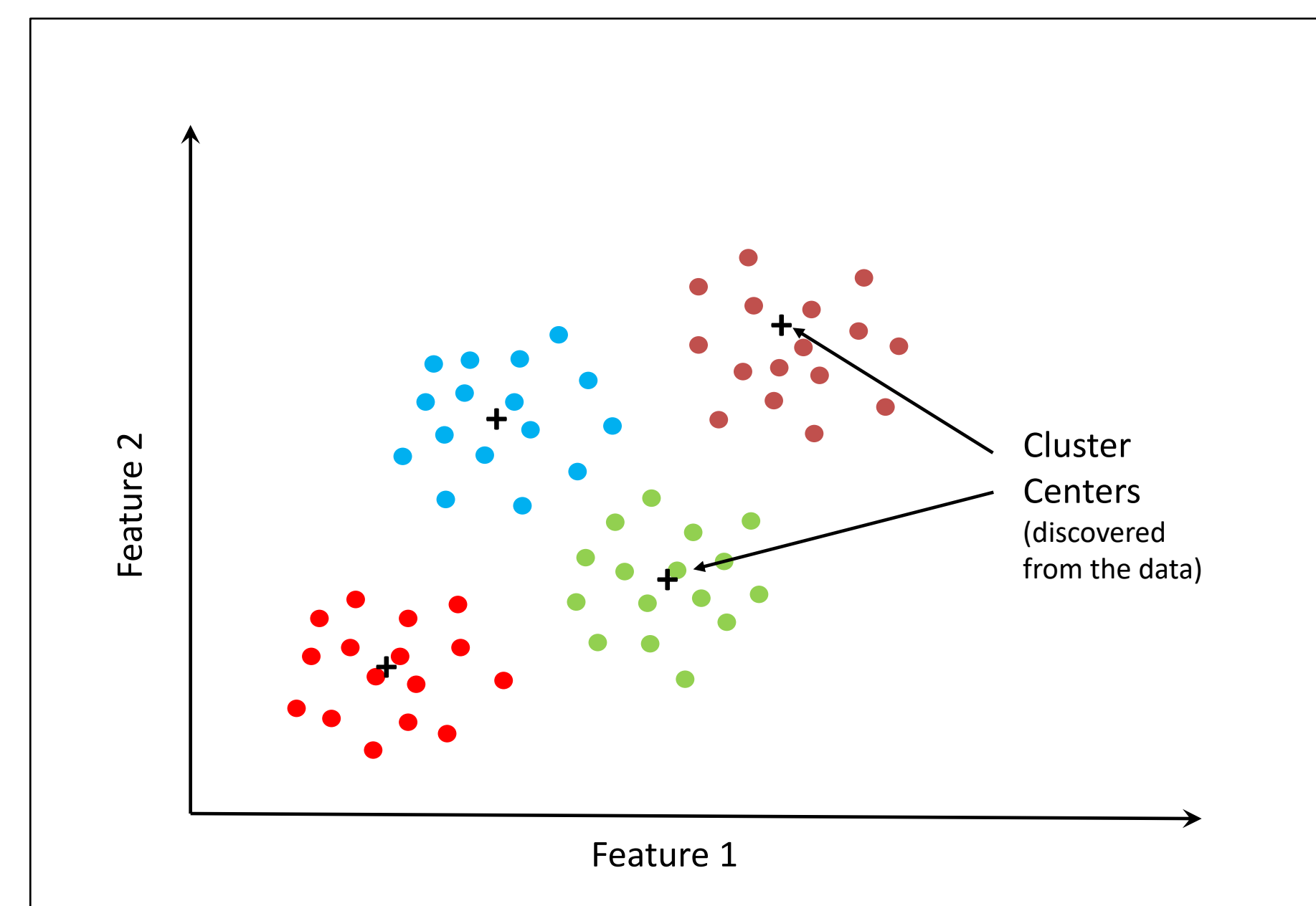


Wave spectrum partitioning (top), and examples of composite wave height and period fields without enforcing spatial and temporal consistency (bottom).

## Approach

The input features to the clustering operation are the component significant wave height, period and direction of each computed partition. The data records comprise the feature values of each partition at each geographical location, at each time step in the wave model simulation. The K-Means clustering algorithm therefore groups partitions with similar wave height, period and direction over the model domain, and in time, and assigns them the same label. These, collected in geographical space and time, represent the identified wave systems.

Since the number of wave systems for each model run is not known a priori, and since this is an input parameter of the K-Means algorithm, the analysis is repeated with a range of K values and the one yielding the highest silhouette coefficient is selected.



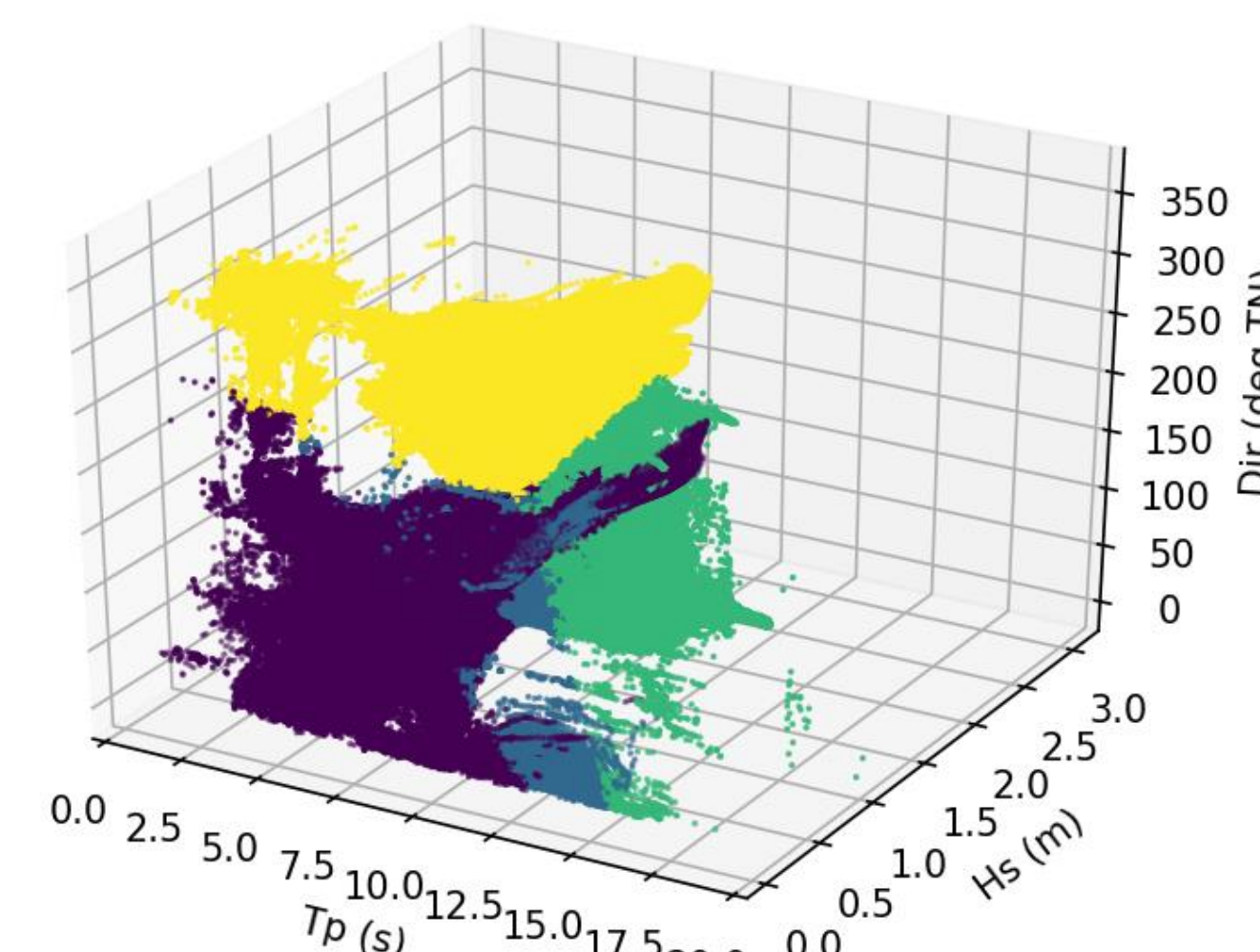
Schematic of the K-Means clustering process, showing data features on the horizontal and vertical axes, and four identified clusters with their centers.

Date	Lat	Lon	HS_PT	TP_PT	DIR_PT	System
t=0	27.0	-79.5	2.1	14.2	60.2	1
t=0	27.1	-79.5	2.2	14.0	59.7	1
t=0	27.2	-79.5	2.0	14.1	58.1	1
t=0	25.0	-82.5	1.2	9.8	295.4	2
t=0	25.1	-82.5	1.4	9.7	293.1	2
t=0	25.2	-82.5	1.1	9.9	294.3	2
t=1	27.0	-79.5	2.2	13.4	55.8	1
t=1	27.1	-79.5	2.4	13.1	55.4	1
t=1	27.2	-79.5	2.1	13.3	54.9	1
t=N	25.2	-82.5	0.5	6.5	275.2	2

Example of the wave partition dataset. Each record represents a unique spatial-temporal partition, which is clustered on the basis of similarities between its features height (HS\_PT), period (TP\_PT) and direction (DIR\_PT).

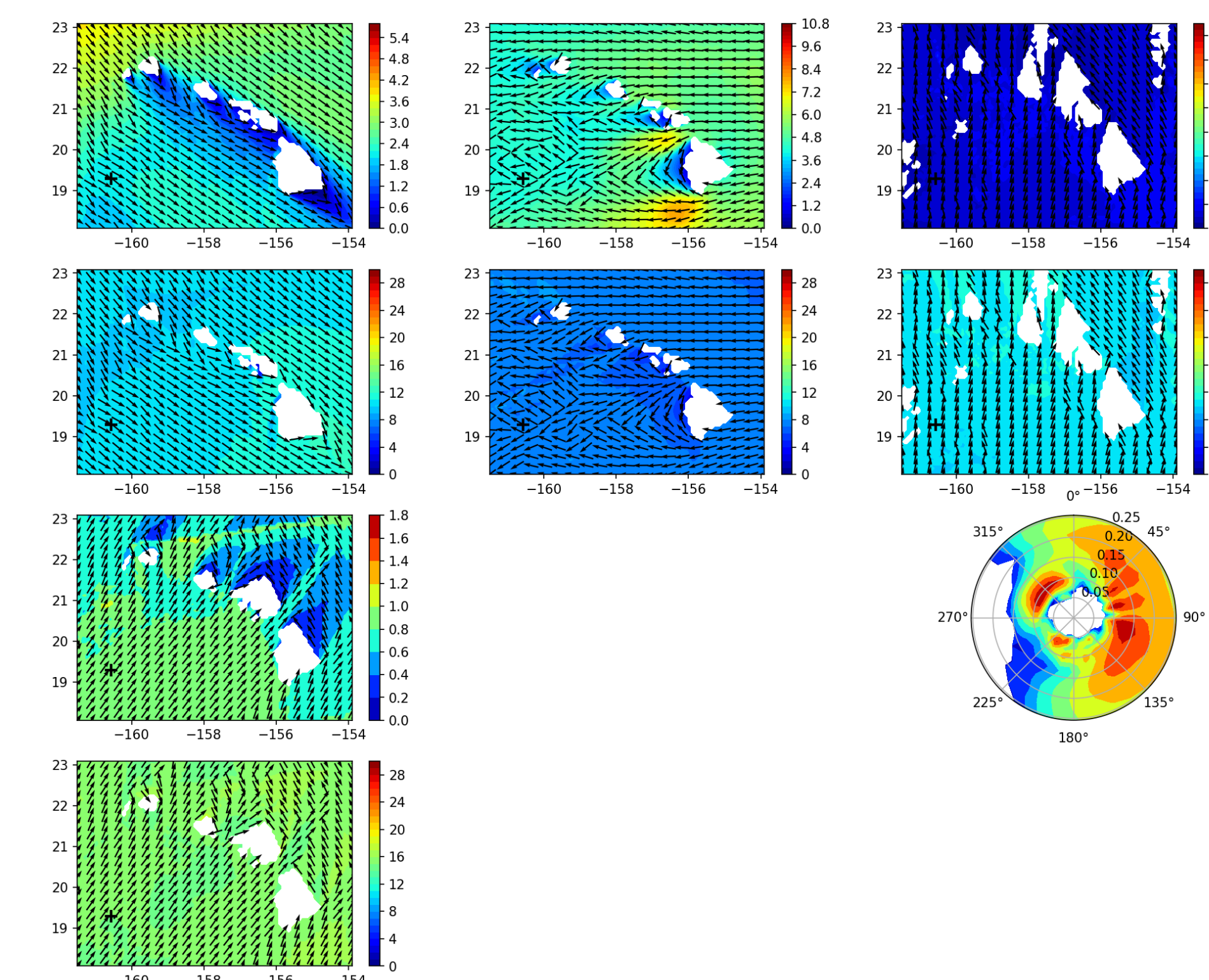
## Three views of the results

The results of this procedure can be viewed in three different ways: first as a scatterplot of the identified clusters in feature space (top), second as mapped onto their geographical coordinates to form coherent wave systems (center), and third as a time series of the different systems at a given location (bottom).

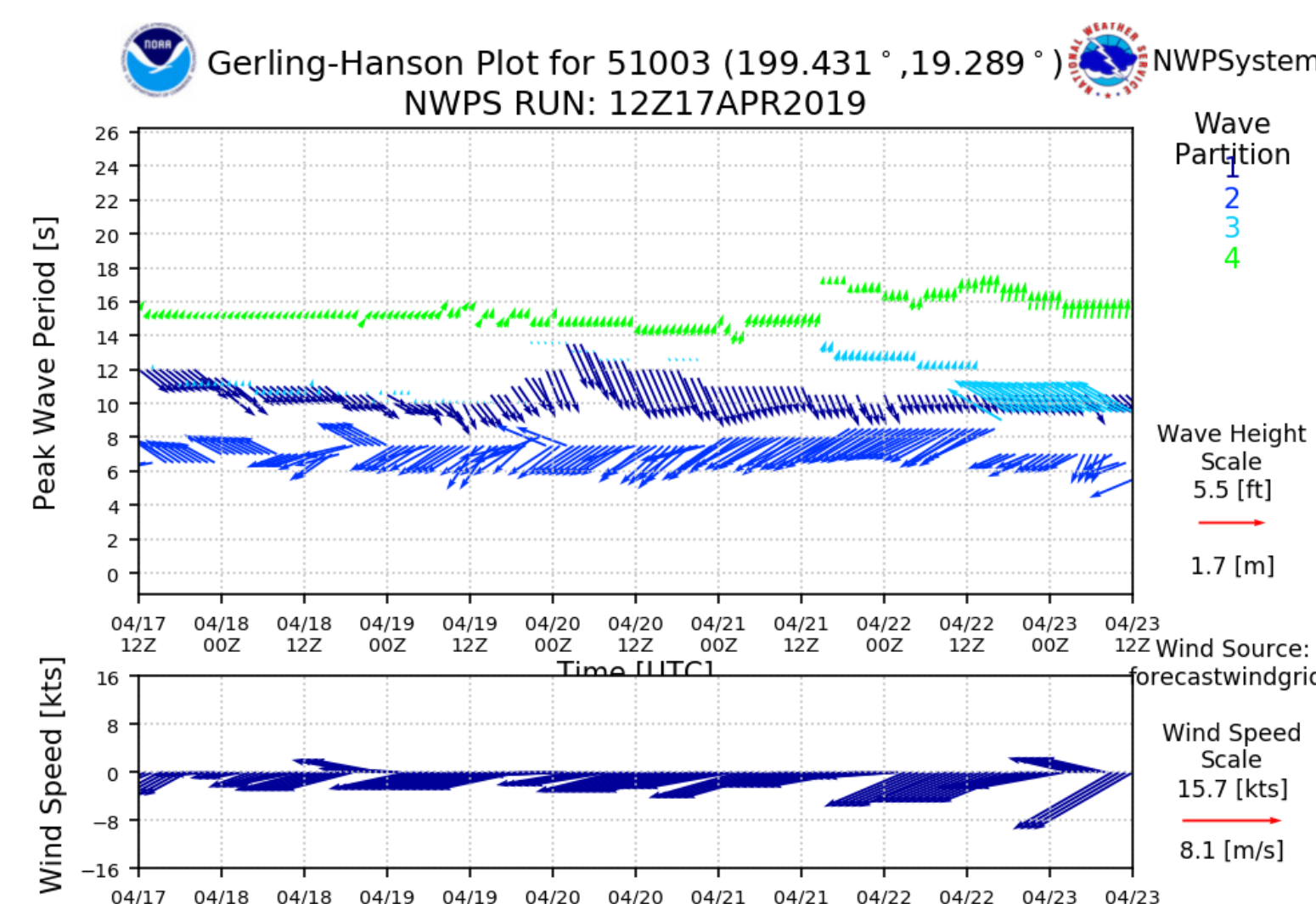


Four identified wave system clusters (colors) for WFO Honolulu in terms of the feature space of wave height (Hs), peak period (Tp) and direction (Dir).

NWPS Wave Systems: Top: Hs (ft) and Dir; Bottom: Tp (s) and Dir  
Hour 36 (00Z19APR2019), SC = 0.61 (0.68/8)  
\*\*\* EXPERIMENTAL - NOT FOR OPERATIONAL USE \*\*\*



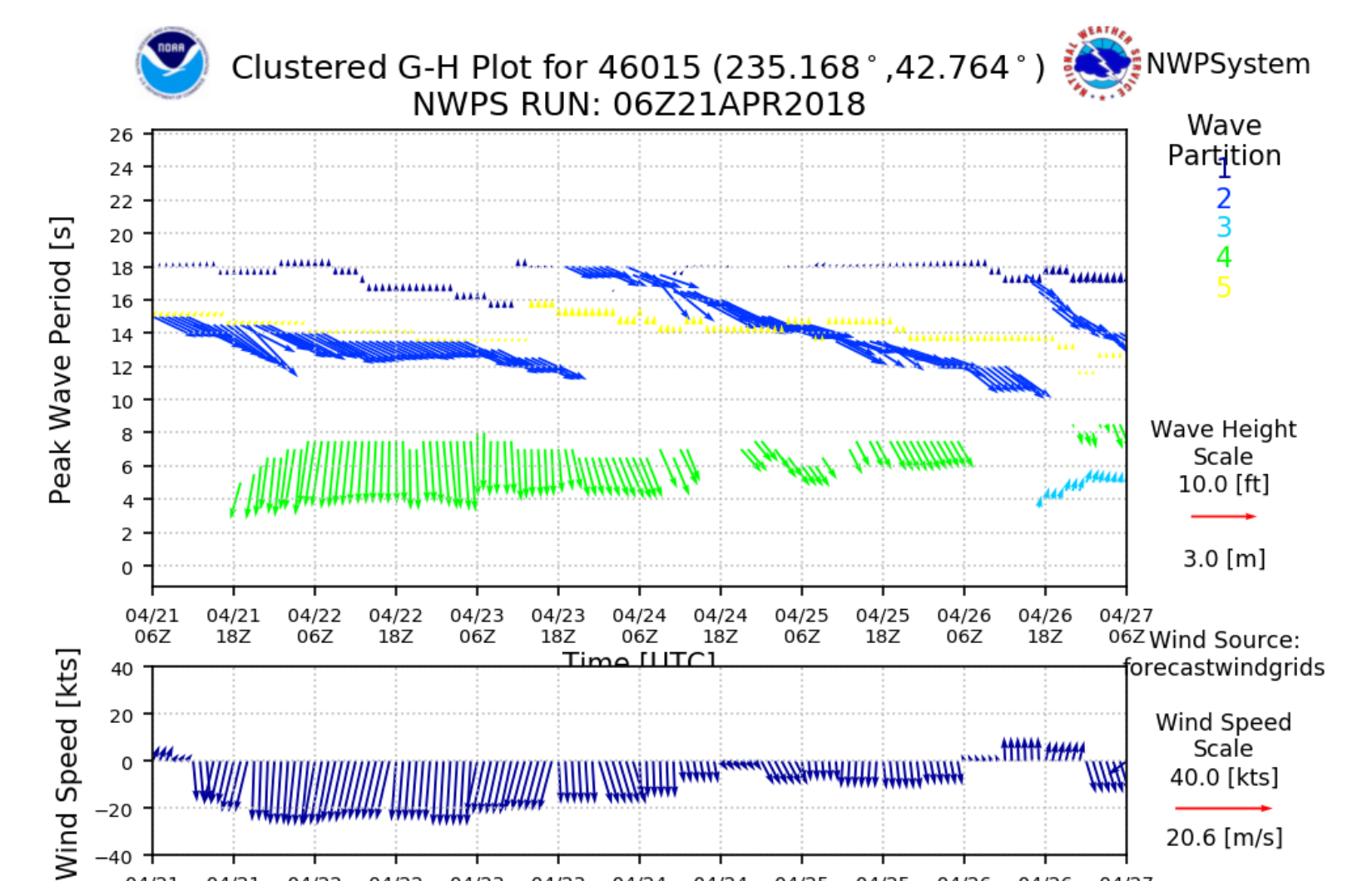
Clustering results in terms of wave fields in geographical space. Four wave systems shown, with Hs in top panels and Tp in bottom panels.



Time series view of wave system clustering results at NDBC 51003.

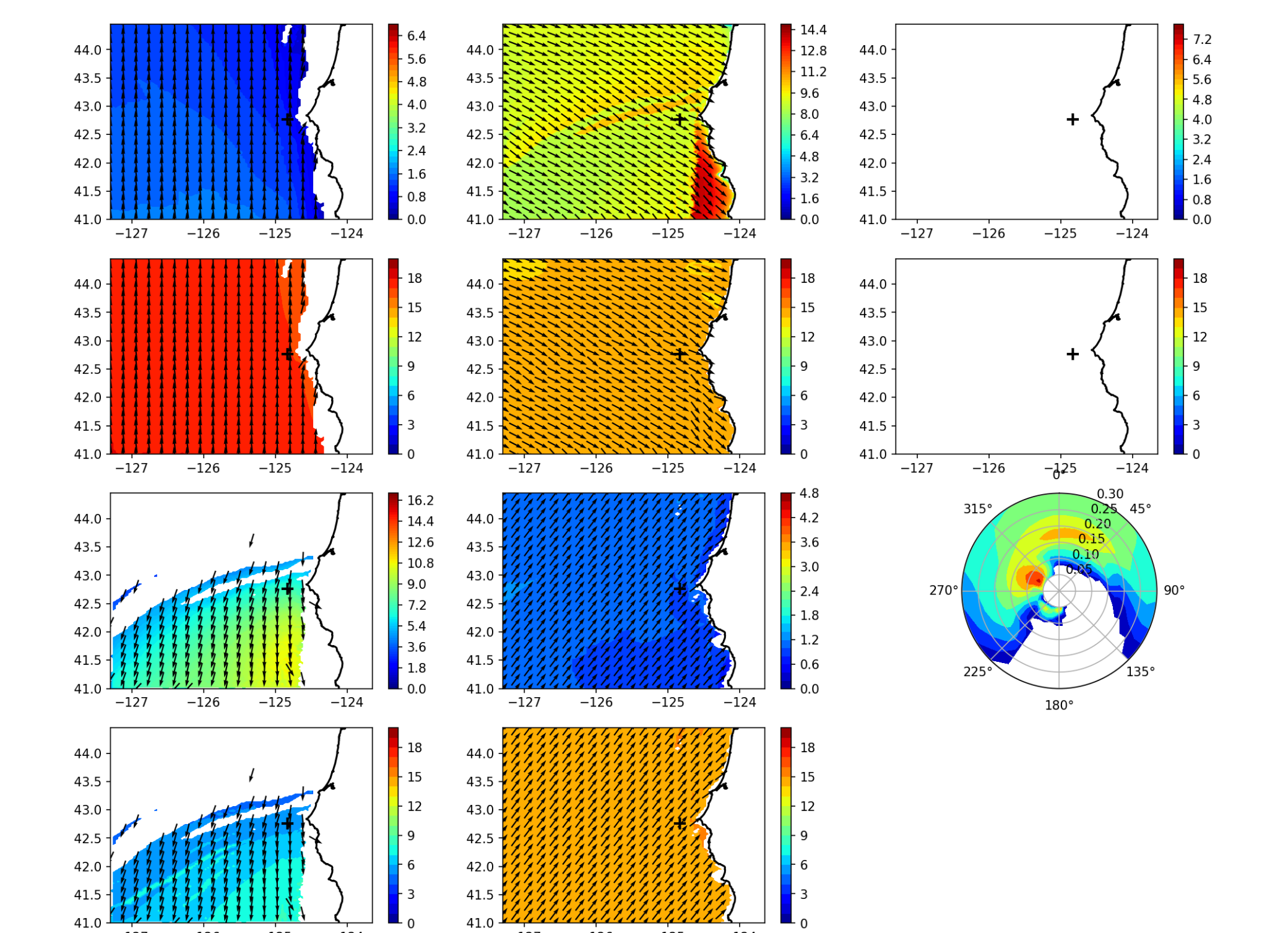
## Remaining Challenges

A few challenges remain with this clustering procedure, such as the identification of young wind sea generated by local winds. The example below shows a time series of the generation of young wind sea at NDBC 46015 offshore of Eureka, CA. The local wind sea is erroneously absent for the first hours of the local wind being active. This problem is related to the shape of the directional wave spectrum in the wave model. Young wind seas can be connected to co-directional swell energy, so that initially no independent, new partitions are identified, and hence no separate system. When the wind sea becomes mature and separates in spectral space, its wave system is identified.



Gerling-Hanson time series plot for WFO Eureka showing delayed wind wave recognition in System 4.

NWPS Wave Systems: Top: Hs (ft) and Dir; Bottom: Tp (s) and Dir  
Hour 15 (21Z21APR2018), SC = 0.69 (0.93/6)  
\*\*\* EXPERIMENTAL - NOT FOR OPERATIONAL USE \*\*\*



Corresponding wave system field plots showing the delayed recognition of wind sea systems (WFO Eureka). Top panels: Hs, Bottom panels: Tp.

## References

Tolman HL, et al (2016) The WAVEWATCH III development group (ww3dg), 2016: User Manual and system documentation of WAVEWATCH III version 5.16., College Park, MD, USA.  
NOAA/NWS/NCEP/MMAB Tech. Note 329:326